'Access is key': US technologist Carl Malamud is on a satyagraha to make knowledge freely available

He is fighting to make Indian Standards for building and manufacturing free, and to prevent scientific knowledge from being locked behind expensive paywalls.

Shoaib Daniyal
Sep 14, 2018 · 07:30 am

 "I have the sum total of human knowledge in my pocket but I use it to watch cat videos instead," goes one popular joke about how the internet is being used today. But cat videos are not the only thing preventing knowledge from spreading through the World Wide Web. More malicious are restrictions on access, imposed by copyright law, faceless bureaucracies and large corporations. It is the second of these that Carl Malamud is battling. For the past three decades, he has been fighting to make public information, well, public. Thanks to Malamud, American citizens can now access securities data for free, along with patents and a large numbers of legal case files. Currently, he is trying to get the state of Georgia to publish its laws and allow the public to access them for free.

The technologist has also been working to make the Indian Standards for building and manufacturing freely available and to prevent scientific knowledge from being locked behind expensive paywalls. Published by the Bureau of Indian Standards, the Indian Standards are required for making everything from milk powder to an electric switch. The "ISI" mark, probably one of the most well known logos in India, certifies that the product displaying it conforms to these standards.

Scroll.in spoke to Malamud about his work, the Indian state's attitude to his advocacy and why he thinks freeing up knowledge is a modern-day satyagraha.

[Boing Boing](#) has called you a rogue archivist.
[Laughs] Yeah, they tagged me with that years ago.

More traditionally, you are described as a public domain activist. What does that mean?
I would not say public domain activist; I would say public domain worker. Activists would simply say this policy is bad. I have tried to publish large databases that should be public and tried to get the government to do the right thing. In the United States, I put the patent database online for the first time. I did it also with the Security and Exchange Commission [which regulates the US securities market]. In India, I started with the Indian Standards. I put all 19,000 on the internet. In fact, we took a thousand of them, retyped, turned them into html format, redid the diagrams, and made them accessible to the visually impaired.

So, you actually spent time cleaning them up?
Yeah, absolutely. And then made them available on the [Internet Archive](#), which is wildly popular with engineering students in India. There are 6,50,000 engineering students in India and every one of them needs to look at the National Building Code but it costs Rs 14,000, which is a lot of money for a book in India. If you are a foreigner trying to buy it, it is Rs 1.4 lakh. They charge 10 times as much from foreigners. That is just nuts. If you believe in [Make in India](#), for example, then you want foreigners to know the rules for safety. It is not just building codes, it is textile machine safety, irrigation, pesticides, firefighter protective gear. A firefighter should be able to pull these in. They are the law, they are notified in the official gazette, many of them are part of regulations. There are 150 products with mandatory certification in India. You can't, for example, sell steel or concrete without BIS certification in India.

Less than 2% of the BIS revenue comes from selling standards. The rest comes from certification, that is the big money. So, we think the standards ought to be available for free. That's one database I have worked on in India. I also have this Hind Swaraj collection up on the Internet Archive that I just loved working on. It began with the works of Gandhi, all 100 volumes. It has around 129 audio files of Gandhiji speaking on All India Radio in 1947-'48. You can listen to his speech, pull out its English version from the archive, the translation of the speech. The footnotes are live. You can read a speech and then click back into the collected works and see the letters he wrote that day, the letters he wrote the next day. You can actually walk through the last year of his life.

Add to that the collected works of Jawaharlal Nehru, the most complete collection online of the works of BR Ambedkar. I have been going to used bookstores collecting the works of S Radhakrishnan. Also of Ram Mohan Roy, the letters of Motilal Nehru, the Doordarshan programme *Bharat Ek Khoj*. The common thread here is that most of these works were funded by the government and most are sold by its publications division. To me, these are meant to be public but many of them are unavailable. So, I go on Amazon and buy, say, a used copy. Then I go to the publications division and buy some more books, and I scan them. I am hoping to give a copy of the Hind Swaraj collection to the Ministry of Information and Broadcasting. It would be wonderful for them to have it on their website.

I have been doing this with other government departments as well. I have been buying their material because, again, I think the works of government should be available to the people. My use is noncommercial, so I am hoping nobody has a problem. Sometimes, the government asserts copyright over such works but I think there is good, sound reasoning for why I am doing this.

Has anyone in India asserted copyright over any of the works you have published so far?
No, I haven't had any problem yet but I may have to face it at some point. You know, Gandhi teaches that you have to educate yourself and believe in what you are doing and suffer the consequences of your actions. So that [legal action] is possible, but I don't think it will happen. We have got another project going in India: mirroring all official gazettes. Not only of the Union government but also of states. I am working with Sushant Sinha and Srinivas Kodali on this.

Now, see, you can't search the Union government's gazette. So, one of the things we have done is that you can search it now. We are adding state gazettes. We are sending law students to ask states for their gazettes since we can't find some of them, like Uttar Pradesh's, online. There are 250 million people in that state but its official gazette doesn't seem to be any place online. So we are asking for CDs and DVDs of these gazettes. What we would like to do is have a full collection of all the gazettes of India. Also, we are finding technical errors. So, we write them letters and ask them to correct these errors. Sushant is working on changing the proprietary fonts to Unicode which means you can cut from and paste to the document. Ideally, we want to classify them, saying this one is about this, that one is about that.

How did you get this idea of being a public domain activist? What set it off?
I wrote half a dozen professional reference books in the 1980s. I was a database expert and I wrote the first book on Ingres [a database management system]. I was self-taught, but in those days everyone was self-taught. I wrote a three-volume series on networking protocols. As part of that I had to buy ISO standards which are very, very expensive. I had to spend thousands of dollars in order to write these books.

More importantly, the internet was built on the telephone network. That meant the International Telecommunication Union standards [published by the United Nations] needed to be public. So, I met the secretary general of the United Nations and said, "I'd like to put the ITU Blue Book [of standards] online for free." He said, "I'd love to make it available on this internet of yours." He

kept calling it that: "this internet of yours". It was a small network in those days. So, we put it online. We soon got a call from the National Science Foundation saying, "Carl, you're using half the bandwidth on the internet." Because everyone was dragging the Blue Book down. We got a fax from the ITU later saying this experiment has been cancelled. Please delete all copies on the internet. That's where I learnt a lesson: if you take a big database and make it available, sometimes you can force change in government. That's why I prefer public worker over public domain activist. Because instead of just writing a letter saying you're doing a bad job, you come walking in with a million people behind you that are now actually using the system. When you do that, politicians get a million people and they go. This is actually doable.

Often, there is stuff that is buried and that could be made available. For example, I once went and met the Archivist of the United States. The National Archives has tens of thousands of videos and you could go in and make a copy. I said to him, "You know that's a clunky system. Could I send in volunteers with DVD duplicators?" He agreed. I put them on YouTube and got 75 million views. They include army training films from the 1940s on how electricity works. I still get notes from students saying, "I never understood what the hell my teacher was saying but this video is great". There's all sorts of amazing stuff that government has put together.

Isn't it paradoxical that public documents, laws and standards are not public. What's the rationale for this?
It isn't that they don't want the stuff to be available, but they haven't kept up with the times. The people at the Bureau of Indian Standards are good people. They just sell standards the way standards are sold all over the world. They're all copyrighted. I've been fighting this battle all over the world. We're in court in the US, we're in court in Europe. So BIS is not unique that way.

I think a lot of people want to do better, they just don't know how. A lot of what I try to do is help make government more efficient.

So, it's more lack of information than malevolence?
I think so. This is different from Julian Assange-Wikileaks model which is to dig out secrets or, in the case of Wikileaks, change the election result. But I'm up for stuff that most people would agree should be public. As a rule, people think that court records should be available in a democracy. Many people think works of government should be available to people in a democracy.

You characterise your work as civil disobedience. What exactly does that mean?
One of the principles of satyagraha is to first petition the government and then do whatever you have to. I went to see Sam Pitroda when he was in [the Manmohan Singh] government and I bought up the Indian standards. He asked, "Why are they copyrighted? Why do they cost so much?" I said I am thinking of posting these on the internet. Sam said, "This is good." I said, "Sam the bureau is going to be very annoyed." He said, "I don't care."

So, I posted them online. A year later I spent $5,000 to renew my subscription. Again, it is a Gandhian thing. I send them a purchase order saying I would love to renew. "And, by the way, these are all your standards, would you like an html?" At which point they cut me off. No more standards for you. They also threatened to invoke copyright so we petitioned the Ministry of Food, Consumer Affairs and Public Distribution. In that, we showed how we had made things better. We had tables, we had listed all the standards that are part of regulation, we had affidavits from distinguished professors. However, the ministry turned us down. So we went to the High Court. October 9 is the day for our oral arguments.

Do you expect the Bureau of Indian Standards to contest your claim?
They already have, they have hired a lawyer. Our position is that [the standards] are the law, developed by a government body, supervised by two cabinet ministers, five state ministers, two members of Parliament and a whole bunch of joint secretaries. The standards are issued in draft, they are developed by professors and engineers appointed by the government who volunteer their time. The public has a chance to comment. They are then noted in the official gazette. Salman [Khurshid], our lawyer argued, this is about the law being available. BIS stood up and said, they [standards] were notified in the official gazette but they weren't published in the official gazette.

What does that mean?
Well, the judge asked, "Young man, what is the difference between notification and publication?" He had no answer, at which point there were 10 minutes of, "Young man, this is an issue of seminal importance. There are hundreds of books out there on what makes a law, I suggest you read one of them before coming to this court." So, the judge sent him back to write a three-page paper on it. At the last oral arguments, we only had five minutes. The BIS said, "Well they are available. You just have to pay money." Again, the judge asked, "Do you really have to pay money to know what the law is?"

But what is driving this? It's not a lot of money for them.
All standards bodies in the world have been doing it this way. I have been challenging it in the US. If it's the law it must be available. But I am getting sued in the US. I have six plaintiffs and four law firms suing me. We lost at the district court level. We now have a favourable Court of Appeals decision. But this is the way it is done. If you think about it, BIS plays in two worlds. They play in the Indian government, but they also play in the world of standards. So, if they say the standards will be available [for free], all their cousins, the American National Standards Institute, the Europeans, they are not going to like it because they sell these things for a lot of money. That's what is driving this.

One of our counter arguments is that not only are these standards important for government regulation, they are important for education. The right to education is there in the Indian Constitution, the right to practise your profession is in the Indian Constitution. BIS standards are about how to practise your profession safely. So, if you believe in these underlying values – the

law must be available, the right to practise your profession, the right to education – then it's a no-brainer that the standards must be public.

Alexandra Elbakyan's Sci-Hub is a 'salt factory on the edge of the ocean of knowledge'. Photo courtesy labiotech.eu

What brought you to India?
It was a couple of things. The Dalai Lama wrote the forward to one of my books, so I was able to go to Dharamshala and present it to him. I finished my first database book in a houseboat in Srinagar, so I knew India. I got to know Sam Pitroda and he invited me to go on the road with him for a week because he was doing a workshop at Sabarmati Ashram on Gandhijis birthday. This is two years ago, in 2016. I went with Sam, travelled around India with him. I went to Mayo College where he gave a speech, so I got invited to give one this year. We went to Rajasthan Central University where he is the chancellor and I got to know folks. A year later, I did another trip with him. Through that, I got to know a lot of people in India. I began coming here on my own, doing projects like the gazettes. Sam gave me a copy of the collected works of Gandhi and I began supplementing that with other things.

Today, scientific journals are totally unavailable. Sci-Hub is a big pirate site [with academic papers for free download]. India is the second largest user of Sci-Hub. It's ironic that knowledge has become colonised. Gandhi did two things: he liberated India but he also decolonised the world. Today, scientists are the new Indigo farmers. They make the raw material. Journals are the railroads that take that raw material and ship them, ironically enough, to a big United Kingdom corporation called Reed Elsevier, which has half the market. They ship back these high-priced finished goods that you have to buy in order to continue farming. Sci-Hub is a salt factory on the edge of the ocean of knowledge. It's a non-licenced factory. But there is a right to

education in the Indian Constitution. So it seems to me that if you believe that knowledge has been colonised, that access to information is a fundamental human right, if you want to solve growing economic inequality, access to information is a precondition.

If you feel that government is not paying enough attention to pollution and global warming, the only answer is to educate the citizenry and own your government. You need to force governments to do the right thing. Unless people are standing up and saying, "I believe global warming is important", governments are never going to do anything about it. So to me, access to knowledge in a democracy is key. If you are going to have that revolution, it's not going to be in the US, it's not going to be in Europe, it has got to start in India. That's why I have begun focusing my attention on India.

I want to come back to the point you raised about the academia. It blows my mind that an industry focused on knowledge would have these incredibly expensive paywalls. I want to know what is driving this remarkable high-priced access to journals and academic books. The academic community allowed private publishing to take over the world when it was just paper copies and the prices were reasonable. It was not a big deal. But then all of a sudden these monopolies came into being. So, there was Reed Elsevier, but also, if you are doing chemicals, the American Chemical Society is where you want to publish. These guys had monopolies and the prices started going up and up and up. They began abusing the monopolies. There are actually a whole bunch of articles that are not under copyright but inside these paywalls and you can't get to them. They have these special packages that are so expensive that even Harvard University is scaling back. Most universities in India do not have full access to the scholarly corpus. You cannot do big data on the entire sum of human knowledge because the corporations will not allow you to do it and that means lives are being lost because there is a lot of valuable research that can be done by taking all the journal articles and doing big data on them: looking for different citation metrics, looking for linguistic analysis over time, looking for meta research on genomics. If you want to find all the articles about genes, you can only do that by having the full corpus and doing a data search. But you can't do that today. Ironically, many of those involved in this are non-profit academics societies. But all of them have become very greedy. The prices have gone through the roof. It is the same thing with standards: they used to be reasonably priced but then it just started going up.

What is the inflection point for this sort of pricing?
It was in the 1980 when the prices really started going through the roof. Ironically enough, just as computers began to go in.

Just when you could actually have had free access?
Yeah.

So again, what drove this?
Greed. Company started getting more and more possessive. We have seen that [high pricing] in academic publishing with journals. They also do these abusive agreements like package deals

where you have to get the whole package even if you don't want all the articles. Many times the bundle pricing is secret: you can't tell your fellow librarian how much you paid for the bundle. Then what will happen is that if you're a researcher and need a particular article that's not available in your bundle, they will charge as much as $5,000 for a single article. There is documentation on this. Then you will have to go back to your funding agency and say, I really need this.

There is this feeling that publishing has to be expensive since you have these layers and layers of executives and corporate types. But at the end of the day journals are edited by professors who aren't paid to do that. And peer review is done by fellow academics. I see this also with the standards bodies. The head of the American National Standards Institute gets $2 million a year and he runs a non-profit. The National Fire Protection Agency, which is suing me, was paying its head $1.5 million a year and when he retired, he got a $4 million payout on top of that. So, they need the money but not nearly as much as they are getting. A lot of it is just greed.

Do you see a silver lining in India with the Delhi University photocopy case?
Yeah, absolutely. I have looked at that. It seems to me that based on the DU case, there is a path to making all scholarly knowledge available in India under the teaching exemption. I'm not exactly sure what that path is. But that's why I am traveling all over. I am giving speeches to the Indian Academy of Sciences, for example, meeting vice chancellors. I am talking to a lot of lawyers who are intellectual property experts. That's why I did four different law schools on this trip, partly because distinguished intellectual property professors were my hosts or I could go talk to them. I am not sure exactly what the path is but you know the teaching exemption is in the Berne Convention. It is in the Indian Copyright Act. I believe there is a path forward that would make scholarly information available to the students of India.

Do pretty clear that Alexandra's is a pirate site. Like I said, it's an unlicensed salt factory on the edge of the ocean of knowledge. Morally, I think what she is doing is incredibly important but I am looking for a legal path to make that information available. I want to be able to stand up and say we are using all the scientific journals of the world legally in India. That's a different thing from what Alexandra did who simply wanted to make information available to people and has succeeded **you think there is a legal path, though? Alexandra Elbakyan, the person who started Sci-Hub, is facing legal action.**
I want to make it clear that I am not going to put another pirate site up. I am not doing that. It is wildly in doing it so. I have great admiration for her bravery but I am doing something different. What I try to do is confront the authorities and figure out how to get government or academica or the publishing world to do the right thing. That's why I call it satyagraha because it really is an attempt to change the way the government works.

Support our journalism by subscribing to Scroll+ here. We welcome your comments

at letters@scroll.in Carl MalamudPublic Doman